Using k-means to find the optimal placement of Hubway stations

Rachel Chmielinski and Rhonda Mak

CS 591 L1

Motivation

Boston is a busy city where many rely on public transportation. Unfortunately, we cannot expect to have all of our trains to run on time — in which case we have to take matters into our own hands if we want to be punctual. You could call a cab, a Lyft, or an Uber instead of waiting. But why not bike to where you need to go instead of contributing to traffic?

Objective

The goal of this research is to find the optimal placement of Hubway bike stations, considering which stations consistently run behind schedule. This will maximize the number of options commuters have available to them, given that a train is running late. Moreover, it will decrease traffic as well as promote eco-friendly transportation.

Methods

Our project utilizes k-means (specifically Lloyd's algorithm, which we used the SciKit KMeans library for) to identify where trains are most often late and place a Hubway station there. We have the following data sets (and the algorithms associated with them):

1. MBTA Green Line stations and their coordinates

2. Hubway stations and their coordinates

3. MBTA Green Line stations and their reliability (a metric that mea sures how many people who have waited longer than the expected time for their train, calculated by MBTA)

Determining the optimal location for each Hubway station based on train reliability:

1. Create a dictionary called lateTrains where keys are stations and values are reliability (rounded down to nearest integer).

2. Create a lateTrainsCoords list, which is a list of coordinates of late trains. For each station x in lateTrains, add the latitude and longitude values corresponding to station x, which we get from our list of station coordinates. We do this n times for each station, where n is the reliability constant for each station x. For example, if lateTrains looks like this:

```
lateTrains = {"Hynes Convention Center": 2,
"Babcock Street": 4}
```

and we know from that Hynes Convention Center is located at (42.0123, -71.456) and that Babcock Street is located at (42.0789, -71.000), then step 2 will create a list that looks like

```
lateTrainsCoords = [ (42.0123, -71.456),
(42.0123, -71.456), (42.0789, -71.000),
(42.0789, -71.000), (42.0789, -71.000),
(42.0789, -71.000) ]
```

We run k-means on lateTrainsCoords. The output will give us an idea of where to place a new Hubway station or where to move an existing one. In our experiment we set k = 40, because that is roughly the number of existing Hubway stations surrounding the Green Line.

Results

The average distance found between existing hubway stations and the optimal location calculated by k-means was 640 feet. This is roughly 1.2x the distance between the BU East stop and the BU Central stop on the B-line of the MBTA Green Line. Walking versus biking this distance could be the difference between being on time and late.

It should be noted that the stations with the most frequent delays are in downtown Boston. It should also be noted that many of these stations have Hubway stations very close. When running our k-means algorithm on just the downtown stops (Science Park, North Station, Haymarket, Government Center, Park Street, Boylston Street, Arlington, Copley, Hynes Convention Center, Kenmore), the average distance found between existing hubway stations and the optimal location was 98 feet. This likely because the density of Hubway stations is concentrated in the downtown area.

Our k-means algorithm also included suburbs outside of Hubway's current regions of service. In the interest of increasing punctuality and reducing traffic across the Boston metro area, perhaps Hubway should consider expanding their regions of service.

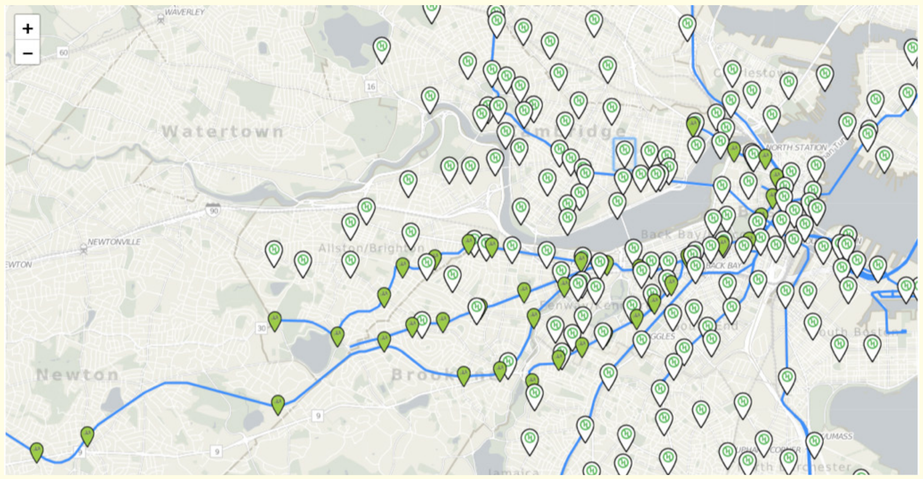

Fig. 1. Plot of current Hubway Stations (white) alongside optimal station placement (green).

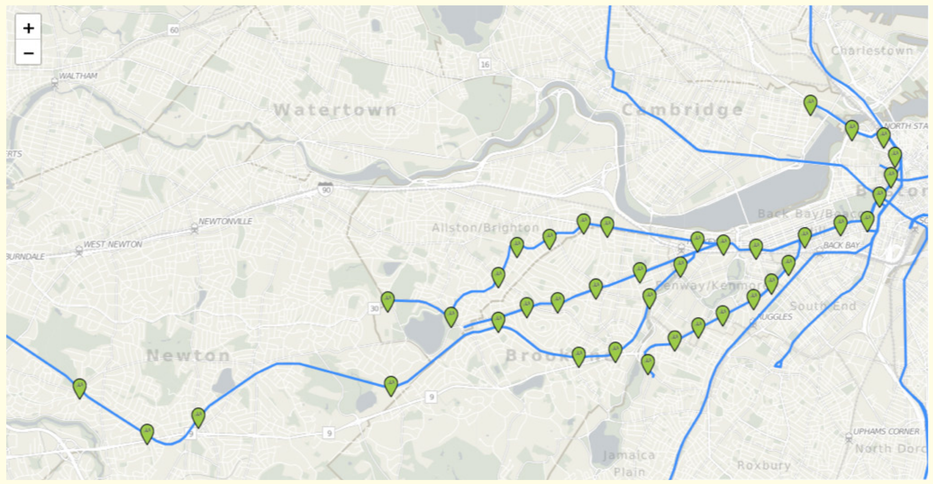

Fig. 2. Plot of optimal station placement (station at Boston Riverside not pictured).

References

Analyze Boston. "Hubway Stations" Department of Innovation and Technology, 2017. Web. 13 March 2018. <http://bostonopendata-boston.opendata.arcgis.com/datasets/ee7474e2a0aa45cbbdfe0b747a5eb032_0>.

MBTA's Fiscal & Management Control Board. "MBTA Back on Track" Massachusetts Department of Transportation, 2017. Web. 13 March 2018. http://www.mbtabackontrack.com/performance/index.html#/download>.